DOCUMENT RESUME

ED 063 324 TM 000 320

AUTHOR Flaugher, Ronald L.
TITLE Testing Practices, Minority Groups, and Higher Education: A Review and Discussion of the Research.
INSTITUTION Educational Testing Service, Princeton, N.J.
REPORT NO RB-70-41
PUB DATE Jun 70
NOTE 36p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Ability Grouping; Admission Criteria; College Admission; Competitive Selection; Disadvantaged Groups; Educational Discrimination; Educationally Disadvantaged; Environmental Influences; Higher Education; *Literature Reviews; *Minority Groups; Objective Tests; Prediction; *Predictive Ability (Testing); Predictive Validity; Racial Differences; Success Factors; *Testing Problems; *Test Validity

ABSTRACT

Some of the controversial issues involved in the use of objective tests by institutions of higher education, as this use affects the selection and attendance by members of minority groups, are reviewed. Admissions committees now rely on the ability of a test to predict students' performance at their institution to guide their selection. However, minority group members have criticized such uses of test scores. Three potential sources of bias against minority groups include: irrelevance of the test content, particularly verbal content, to their culture and background; discriminatory administration of the testing program; and discriminatory use of the test results. Research investigating the comparative performance of minority and majority group members, the predictive validity of tests, and the influence of the testing environment on performance is reviewed. Efforts to isolate culturally biased test items have been unsuccessful. Evidence indicates that minority group members tend to score less well on most tests; however, tests seem to validly predict academic success regardless of the student's background. In addition, the physical and psychological atmosphere in which the test is administered seems to have a significant influence on performance. Suggestions for supplementary research are delineated. (PR)

ED 063324

RESEARCH BULLETIN

RB-70-41

TESTING PRACTICES, MINORITY GROUPS, AND HIGHER EDUCATION: A REVIEW AND DISCUSSION OF THE RESEARCH

Ronald L. Flaugher

This Bulletin is a draft for interoffice circulation. Corrections and suggestions for revision are solicited. The Bulletin should not be cited as a reference without the specific permission of the author. It is automatically superseded upon formal publication of the material.

Educational Testing Service
Princeton, New Jersey
June 1970

TM 000 320

TESTING PRACTICES, MINORITY GROUPS, AND HIGHER EDUCATION:

A REVIEW AND DISCUSSION OF THE RESEARCH

Abstract

This paper reviews some of the specific issues that underlie the controversy concerning the use of objective tests by institutions of higher education, as this use affects the selection and attendance by members of minority groups, or those persons designated by the term "disadvantaged." Following a discussion of the issues, a review of the research literature attempts to reveal what is known about each of them; this is followed by some suggestions for future research efforts.

TESTING PRACTICES, MINORITY GROUPS, AND HIGHER EDUCATION: A REVIEW AND DISCUSSION OF THE RESEARCH[1,2]

Ronald L. Flaugher

Educational Testing Service

This paper reviews some of the specific issues that underlie the controversy concerning the use of objective tests by institutions of higher education, as this use affects the selection and attendance by members of minority groups, or those persons often designated by the term "disadvantaged." Following a discussion of the issues, a review of the research literature will attempt to reveal what is known about each of them; this in turn will be followed by some suggestions for areas in which future research efforts might be directed.

Perhaps it will be most accurate to define the population of interest as do Kendrick and Thomas (1970) who adopt the term "disadvantaged" in spite of the increasing objections to it, but define the term carefully as "members of groups that have historically been underrepresented in higher education and which, as groups, are clearly below national averages on economic and educational indices." The greatest portion of this group to be considered is black, but it includes other minorities, such as Puerto Rican, Mexican-American, and American Indian, and lower SES groups of any ethnic origin.

This paper's concern will be the role of testing in altering this underrepresentation. Various authors quoted here have used various terminology, it should be noted, but the generality of the definition used here should prevent confusion.

There may be objections raised to the treatment of all of these "out-groups" as interchangeable. This paper takes the position that

until more research has been completed and our understanding is greatly increased, more elaborate distinctions are probably pretentious. Of course, when relevant distinctions are made in the research, they will be noted; fundamentally, however, we will define our interest as testing practices in higher education as they apply to this underrepresented "out-group."

By the term "testing practices" we shall be referring to the typical situation in which applicants, in order to gain entry to an opportunity for higher education, must submit scores based on an objective examination. This is usually a group-administered test form, and the test takers are usually required to fill in blank spaces on an answer sheet constructed so that it may be scored mechanically. The resulting test scores of the applicants are taken into consideration, along with other kinds of information, when admissions decisions are made by the receiving institution.

As an index of the amount of underrepresentation of this outgroup in our educational system, Time magazine (1970) provides the following statistics for black Americans. In grade school, only 58% of black school children complete the eighth grade, as against 73% of their white classmates, and about 40% of black teenagers finish high school compared with 62% of whites. In college, black enrollment has almost doubled since 1964, but the relative black total has barely changed: only 6.4% of U. S. undergraduates are black, compared with 5% in 1964; they number 434,000, almost half attending black colleges, mainly in the South; at major integrated universities, perhaps 3 out of 100 students are black. In graduate school, blacks account for an estimated 1% of doctoral candidates, most of them in education, and constitute less than 3% of law students and 3% of medical students.

A Detailed Look at the Nature of the Examination

Before exploring the specific objections that have arisen concerning the use of objective tests in selection, it is appropriate to review the intent and nature of the typical examination as employed in institutions of higher education.

First, the test is administered in an attempt to predict performance at the particular institution. It is made up of a sampling of tasks which tap those aptitudes or bodies of information that have been judged, either statistically or by some other means, to be required for successful completion of the educational goal. Furthermore, the assumption is that the more of the attributes possessed by the applicant, the more likely he is to succeed. Since he is the more likely to succeed, he is judged to be more deserving of the selection, since his inclusion in the student body will lead to optimal use of the resources of the institution which accepts him.

There are, of course, differences between the test and the college experience. The college is a long-term experience, for example, and although speed may be called for frequently, there is not the overriding time pressure typically associated with the test situation. Further, the test is "faceless" and impersonal, whereas supposedly the college is not. The test is taken in a highly structured situation involving only very formal relations between supervisor and students. While the college experience may require highly efficient study habits and in general the adoption of unfamiliar value systems and habits, it nevertheless provides numerous opportunities for fairly free interaction with other persons. It is distinctly possible that any particular person could perform well in one of these situations and poorly in the other, making his test scores unrepresentative, that is, invalid.

Specific Criticisms of Testing

Let us turn now to the specific criticisms leveled against tests and testing practices by minority group members or those who share their particular concerns. Not all of these accusations are put forth by all critics of tests, and indeed, in some instances the specific positions represent contradictions of one another. They will simply be documented here, however, as a preliminary to reviewing the research literature for what may be revealed about these positions.

Perhaps the most commonly heard complaint against most tests is that the content is "irrelevant" to the culture and background of the minority group member. There is an assumption that certain facets of the majority culture are not accessible to the minority group member and that the test is unfair insofar as it requires knowledge of them. The minorities have not had a hand in determining the content of the test, nor a chance to absorb the majority culture on which that test is based and by which it was produced. One particular target of criticism is often the "verbal" content of the test on the grounds that this particular type of verbal experience is not as important for minority cultures as it is for the white majority; therefore, emphasis on this skill in tests is effectively discriminatory.

We have defined the test and its cultural basis and one of the criticisms leveled at testing by minority group members. By doing so we can perceive two distinct and conflicting positions. The one states that the content of the test is determined to the extent possible by the requirements of the task being predicted; verbal and reasoning skills, for example, are contained in the test because they are required in the curriculum.

Regardless of which persons possess this aptitude or information, that is what is to be measured, and those possessing it are to be admitted for pursuit of that curriculum. The tester's attitude is that those who were turned away would not have succeeded, at least to the same degree as those who were accepted, and therefore their rejection is quite fair.

The opposing position, on the other hand, takes the attitude that the test fails to measure many attributes that are related to the success on the task; in other words, it is not valid in any comprehensive sense and, for this reason, when used for selection it is unfair. If the test had not been used, those who were turned away would have been accepted, and furthermore would have succeeded.

An additional criticism of tests is that, apart from the test content itself, the actual examination procedure may have an unfair influence on minority group performance. Sattler (1970, p. 144) notes the paucity of research on the influence of racial factors, but states that:

> Numerous writers have either concluded or suggested that this variable may play an important role in the intelligence test situation (Anastasi, 1959; Anastasi and Foley, 1949; Blackwood, 1927; Brown, 1944; Garth, 1922-23; Hilgard, 1957; Journal of Social Issues, 1964; Klineberg, 1935, 1944; Pettigrew, 1964; Pressey and Teter, 1919; Strong, 1913). These writers have suggested that racial examiner-examinee differences, primarily between white examiners and Negro examinees, may lead to such examinee behaviors as fear and suspicion, verbal constriction, strained and unnatural reactions, the assuming of a facade of stupidity in order

> to avoid appearing 'uppity,' and scoring low in order to avoid personal threat. Not only are rapport difficulties postulated, but Pettigrew (1964) also suggested that Negroes may view the test situation itself differently from whites: Negroes may perceive the test situation as a means for white persons to get ahead in society, but not as a means for themselves to get ahead. Many of these behaviors, patterns, and perceptions are likely to exist, and are important phenomena in their own right; it is still not known to what extent they affect the examinees' scores.

Still another criticism of testing asserts that there are aptitudes in minority groups that are not tapped by the traditional test content. It is not that minorities are deficient; rather, they are different, and their different aptitudes need to be recognized in the educational system, even if the system has to be changed to permit this. In other words, tests are predictive of success as the educational process now exists; however, it is imperative that the educational process be changed to take account of those abilities which now exist in minority group members but are not being utilized in educational processes. Some of the research evidence bearing on this important question will be considered later in this paper.

Three Potential Sources of Unfairness

As a final note for this section of the paper it might be helpful to make a distinction among the various possible sources of unfairness that exist within educational testing practices. There are at least three of these discernible: The first and by far the most commonly referred to

is that of the test content. There is a widely held belief that the kinds of tests, or the kinds of questions asked within the test, are biased against minority groups, causing them to perform poorly in ways that are not valid. Second, the test program itself may be conducted in such a way that the result is discriminatory. For example, information essential to registering for and taking the test may not be disseminated in a form that makes it available to minority groups, or conditions may be allowed to exist in the test administration itself which are intimidating. Third, discriminatory practices may exist in the use to which test results are put, such as requiring high verbal test scores to qualify for a job which in fact does not depend upon verbal skills, or requiring certain aptitude levels for graduation from a program rather than using the aptitude measure to select or to predict success upon entering.

Test content and test environment have been subjects of some research, and these topics will be discussed in the following section. Test use, however, is seldom regarded as a subject for at least the ordinary kinds of research effort. Unfairness from any source, however, can be the weak link in an otherwise strong chain and misguided use of test results can be a very serious defect in a testing program.

Review of the Research

Minority and Majority Test Performance

In the great majority of research studies reported in the literature, members of minority groups have done less well in test performance than have the members of the majority groups. Jensen (1969a, p. 81) provides an up-to-date review:

It is a subject with a now vast literature which has been quite recently reviewed by Dreger and Miller (1960, 1968) and by Shuey (1966), whose 578 page review is the most comprehensive, covering 382 studies. The basic data are well known; on the average, Negroes test about one standard deviation (15 IQ points) below the average of the white population in IQ, and this finding is fairly uniform across the 81 different tests of intellectual ability used in the studies reviewed by Shuey. This magnitude of difference gives a median overlap of 15 percent, meaning that 15 percent of the Negro population exceed the white average. In terms of proportion of variance, if the numbers of Negroes and whites were equal, the differences between racial groups would account for 23 percent of the total variance, but--an important point--the differences within groups would account for 77 percent of the total variance. When gross socioeconomic level is controlled, the average difference reduces to about 11 points (Shuey, 1966, p. 519), which, it should be recalled, is about the same spread as the average difference between siblings in the same family.

Most of these studies have been concerned with Negroes, but although it is less well documented for other groups this finding appears to be true for them as well, with the exception of orientals. Coleman (1966), for example, found lower scores for Mexican-Americans, Puerto Ricans, and American Indians.

This general rule applies to measures of aptitude, intelligence, achievement and many of the more obscure measures. There are, however, a few notable exceptions to which we will refer later in this paper.

Contrary to the popular impression, attempts to document the reasons for this discrepancy indicate that the verbal component does not appear to be the cause. Tenopyr (1967, p. 2) reviewed several of these studies:

> The greatest racial differences on tests may not be associated with the verbal abilities, but, instead, are more likely to be attributable to abilities in the non-language areas. Fifer (1965) found Negro children to score relatively higher on a verbal test than on reasoning, numerical and space tests. Studies by Davidson, Gibby, McNeil, Segal, & Silverman (1950) and by De Stephens (1953) indicated that the performance subtests of the Wechsler contribute somewhat more to ethnic difference than do the verbal subtests. For low socioeconomic status children, Higgins and Sivers (1958) found significant differences between Colored Raven Progressive Matrices means for Caucasians and Negroes, but did not find significant differences between Stanford-Binet means for the same groups. Vernon (1965), in comparing test performance of native British children and Negro Jamaicans, found the smallest differences between the two groups were for verbal tests and that largest differences were for spatial tests such as the Kohs Blocks. Moore and MacNaughton (1966) in a study of job applicants

at a southern petroleum refinery, found that the use of essentially non-verbal, spatial type tests resulted in somewhat greater score differences between whites and Negroes than did the use of more verbal tests.

Tenopyr (1967, pp. 6-7) then conducted a study of her own:

> For 500 machine-shop trainee job applicants, including 187 Anglos, 283 Negroes, and 30 Spanish-Americans, it was found that, with a socioeconomic status measure controlled, there were, for Anglos and Negroes, highly significant differences between means on three Employee Aptitude Survey Tests. The largest difference between means for the two groups was associated with the spatial, not the verbal test. This finding provides further evidence that Negro job applicants might be put to as great a disadvantage or an even greater disadvantage if verbal employment tests were replaced with spatially-oriented "culture-fair" tests.

In addition, since the publication of Tenopyr's study, the Fifer (1965) results have been essentially replicated on another population with essentially the same results (Stodolsky & Lesser, 1967).

If, in fact, cultural considerations were of paramount importance, then it should be possible to find some particular test items on which blacks function better (or worse) than would be expected from their performance on the test as a whole. Efforts thus far to find such items have not been very successful. Cleary and Hilton (1968) examined the interaction of individual PSAT items with race and concluded that the interactions contributed minimal

percentages of the total variance of an observation and that, "given the stated definition of bias, the PSAT for practical purposes is not biased for the group studied." Coffman (1965) reports a study which found that two out of 42 SAT questions showed differential difficulties, but the difference was that "the items were easier for the Midwestern sample and both involved content with a rural flavor" (p. 87). Cowell (1969) compared black and white performance on items in the Admission Test for Graduate Study in Business and did find that items involving percentages were relatively more difficult for a group of 110 black examinees.

Validity Studies

Lower test scores for a particular minority group are not in themselves evidence of unfair testing practices. Assuming a fair, unbiased criterion, ultimate conclusions about the question of test bias must rest on evidence concerning the validity of the particular tests in predicting the criterion. Regardless of the score distributions of any subgroups, if the success of the members of these subgroups is predicted equally well using the prediction procedures appropriate for the entire group, then the practice is not discriminatory. Kendrick and Thomas (1970, pp. 162-163) very recently reviewed the existing research evidence concerning the validity of college admissions tests and the possibility of their differential predictability. Their summary is quoted at length here, being the best and most recent summary of what is known about this important question:

> Studies conducted by Boney (1966), Hills, Klock, and Lewis (1963), Roberts (1962), and Stanley and Porter (1967) give evidence that the Scholastic Aptitude Test (SAT) of the College Entrance Examination Board is as valid for predicting

grades of students in predominantly black colleges as for predicting the college grades of white students. Further, when SAT scores were used in combination with school rank, similar predictive validities have been found between black and white students (Olsen, 1957; Roberts, 1964). The possible bias of the SAT in predicting college grades of black students at integrated colleges was investigated by Cleary (1968). She concluded that there were no significant differences in prediction for black and white from the two Eastern colleges selected for the study. Although there was a difference in the regression lines for black and white students at a third college (located in the Southwest), it was a matter of over-prediction of black students' college grades by the use of the white or common regression lines. Morgan (1968) indicated the utility of the SAT mathematics score for identifying "calculated risk" students. Munday (1965) found that the American College Testing Program (ACTP) battery was as useful for predicting the grades of socially disadvantaged students as it has been found in predicting the grades for other students. A few studies have produced some evidence that perhaps the relative utility of high school grades as predictors of college success for students from socially and economically excluded ethnic groups should be reappraised (Thomas and Stanley, 1969). Munday (1965) employing five separate criteria (college English average, college social studies average, college mathematics average, college science average, and overall

college average), found the multiple R derived from optimally weighting four high school grades in each category was lower than the multiple R derived from the optimal weighting of the four ACT tests. McKelpin (1965) found the SAT-V for males correlated higher with first semester average grades for entering freshmen than high school grades did with the same criterion at a predominantly black college in Durham, North Carolina. No substantial difference in the predictive validities of the two pre-admissions indices were noted in the case of black female students. Reexamination of Cleary's data (1968) mentioned earlier, revealed that for blacks in one of the integrated colleges SAT-V and SAT-M correlated higher with college grade point average than did high school rank. Such relative superiority of test scores over high school grades have been noted in the data provided in studies by Funches (1967), Perlberg (1967), and Peterson (1968).

There are currently under way a number of additional studies of this same problem, but the evidence seems to be accumulating that the validity of this type of instrument is not radically different for the minority group members. Furthermore, just as the actual conduct of substantive research has yielded results which differ from the "known fact" that verbal material is the offending component of tests, a number of studies seem to indicate a different situation than had been anticipated regarding the differential validity of these tests for minority groups. Cleary's (1968) study was cited above in which she found that the difference in prediction which

occurred actually was unduly favoring the selection of black students; in other words, their performance was overpredicted by the test statistics.

It happens that this particular finding is not isolated, particularly if we include validation studies from industrial uses of testing. Tenopyr (1967, p. 15) found the following:

> With respect to all criteria, the tests were found to be equally valid for Anglos and Negroes. Assuming the unfair discrimination results whenever, for any subgroup of the population, the criterion scores predicted from test results are consistently higher or lower than actual criterion scores, it was found, relative to six of the ten criteria, that the use of the common Anglo-Negro regression line would result in unfair discrimination. This discrimination, however, would favor, not penalize, the Negroes. In each of these six instances, the use of the regression line based on Anglos alone would again favor, not penalize, the Negroes.

Similar findings were reported by Grant and Bray (1970) from a study of telephone company installation and repair occupations. Using aptitude tests to predict success in training, they found about equal prediction for minority and nonminority trainees, but that the use of a common regression line "biases the use of the tests for making predictions somewhat in favor of minority group applicants" (p. 14).

In addition, a study of medical technicians found a number of instances where aptitude tests, had they been used according to prediction procedures developed for whites, would have overpredicted a job knowledge criterion

for Negro incumbents (Campbell, Pike, & Flaugher, 1969). These authors state the dilemma this way: "The results of the present study. . .not only fail to expose the type of bias which would ordinarily be predicted, but in fact, present some evidence for the existence of what might be called a reverse unfairness. To the authors' knowledge, no hypothesis or theory exists to explain this phenomenon" (p. 7).

Since that writing, however, a possible explanation of these results has been offered by Linn and Werts (1970). They have shown that the over-prediction in many cases may be attributable to one or both of two possible weaknesses in the empirical study: (1) lack of appropriate correction for the reliability of the predictor, or (2) omission of any variable from the regression equation that is related to the criterion on which there are preexisting group differences. Both of these would operate in the direction of creating an overprediction of the lower scoring group, and are probably accounting for at least part of the findings described above.

Rock (1970), however, has offered a different sort of explanation, at least as these findings are revealed in academic prediction, in terms of differing motivation for achieving the criterion of success. In several studies of moderating influences on the prediction of over- and under-achievers (Flaugher & Rock, 1969; Klein, Rock & Evans, 1968; Rock, 1968), it was found that those who were underpredicted (and thus for whom the selection test might be considered unfair) were from backgrounds likely to be characterized by higher than usual motivation toward achievement in college. This characteristic makes them likely to utilize to the maximum what aptitudes they possess, so that their predicted grades are lower than the ones they eventually attain in fact.

One might make a very tentative guess that the same influence is reflected in the results in the industrial studies cited above, in favor of minorities. Minority group members, up to the present time at any rate, have had reason to be less encouraged to try for the traditional goals that are valued as legitimate criteria by the majority group. This would account for the overprediction by the measures of aptitude.

Much more careful research in this area is needed before such explanations can be properly evaluated, of course, and for that matter there need not be just a single process operating in each of these settings. In the following section, the evidence will be examined concerning lowered performance on the predictors, i.e., the tests, rather than the criterion.

Research on the Testing Environment

Concerning the atmosphere in which objective examinations are conducted, we have mentioned this as one of the criticisms of testing as used for minority groups. It is a familiar statement that some people "clutch up" on an examination, while others excel in such a situation. Further, there is among minority groups a sense of a difficult hurdle to overcome, and this is likely to arouse anxieties, particularly for a minority group member who might see this as a method of gaining access to the benefits of the establishment.

Sattler (1970, p. 144) has reviewed the literature concerning such influences on test performance:

> Little is known about the effects of the examiners' race on scores obtained on group administered intelligence tests. Shuey (1966), in a comprehensive review, compared the intelligence test results obtained separately by white

examiners and by Negro examiners in studies using both individual and group assessment procedures, and concluded that white examiners did not adversely affect the IQ of Negro examinees.

However, a series of studies by Katz and his associates (reviewed in Katz, 1970), studying Negro college students, has found an interesting interaction of environmental influences on a series of cognitive tasks. By using both black and white test administrators, and in addition manipulating the information given to the examinees concerning the comparison groups against which they were competing, Katz has been able to show that performance changes systematically as a function of these variables.

Katz interprets his results to mean that the normative group, defined by telling the examinee that his test results will be compared with either black or white performance, determines the examinee's perceived probability of successful performance. Specifically, if black norms are used, the probability of success is viewed as high, while white norms are viewed as more difficult and the probability of success as low. With high probability of success, Katz's results show that the presence of a white test administrator is optimal, while in low probability of success conditions the use of a black administrator is likely to yield better test performance. Katz has also been able to show that apparently other variables, such as past history of successful competition on the part of the examinee, also will influence the perceived probability of success in a given test situation and will thereby alter the test performance.

A wide range of variables which might influence minority group performance in the testing situation evidently remain unresearched. It is entirely conceivable that such factors as guessing instructions and speededness of the examination could exert detectable influences which would interact with race. Although a recent report found no differential advantages by race or SES when additional practice and lenient time limits were permitted for high school students (Dubin, Osburn, & Winick, 1969), these as well as other variables deserve careful additional study, particularly as they interact or result in cumulative effects on test performance.

These studies have frequently employed rather limited types of cognitive tasks, such as arithmetic or digit-symbol tests, which are not ideally representative of aptitude in general; further, they have been performed in individual, or small group settings, rather than large group settings more typical of admissions test administration, and thus the effects of the race of the examiner, for example, are very likely to be at a maximum. Although the actual research remains to be done, there is good reason to believe that perceived probability of success, in particular, would have an influence in virtually any testing setting. In general, the implications for group administrations are clear: the test administration environment can have an influence on test performance; there is potentially a very real source of differential, and hence inequitable, influence on test scores.

Differential Patterns of Ability

It was mentioned above that there are some notable exceptions to the general finding that minority group members tend to score lower on tests. Jensen has conducted a number of studies with children, involving black,

Mexican-American, and Caucasian, and has found that when IQ test scores are equated for groups of lower-class and middle-class children, the lower-class children obtain higher scores on certain tasks of "direct" learning ability: "serial and paired associate rote learning. . .selective trial and error learning, free recall. . .and digit span" (Jensen, 1969b). An isolated study by Iscoe and Pierce-Jones (1964) also found divergent thinking to be measurable in lower-class Negro children in greater amounts than white. Semler and Iscoe (1966) compared the performance of Negro and white children on four conditions of paired-associate learning tasks; they also obtained WISC data on the children who ranged in age from five to nine years. Although significant racial differences were present on the WISC, they were not found in the paired-associate learning.

Lesser, Fifer and Clark (1964) studied four ethnic groups and further divided them by SES. They found that ability level depended upon SES, but that pattern depended upon ethnic identity: Oriental, Jewish, Puerto Rican, or Negro. Lending very strong support to the stability and viability of these findings is the fact that they were essentially replicated in another city (Stodolsky & Lesser, 1967). In general they found that while mean differences favored the majority, the amount of overlap was great, but that ethnic identity determined the particular pattern of relative strengths and weaknesses.

The research findings on this topic are still quite sketchy, and some disagreement exists concerning their proper interpretation (Humphreys & Dachler, 1969a,b; Jensen, 1969c), but if they continue to hold up under additional studies, then the implications are very great for institutions of higher education. If indeed there are identifiable patterns of abilities

within minority groups which differ from those which have been traditionally associated with success in higher education, then the problem of under-representation is not likely to be solved by more vigorous searches for traditional talents. Kendrick (1967) has pointed out that colleges will remain segregated racially if they confine their efforts to discovering talented black students resembling the white students already enrolled.

The solution appears to be necessarily with the institution, and a number of the investigators whose work is cited here have spoken eloquently on the topic:

> No effort to add to knowledge about social-class and ethnic-group effects upon mental ability will have tangible or socially useful educational outcomes unless accompanied by simultaneous, coordinated efforts to develop curricula, train teachers, modify social organization, and improve methods for establishing public policies regarding the schools. Each of the many educational efforts which affect children from culturally diverse groups--issues of measurement, curriculum, teacher training, school organization, and so forth--has remained almost entirely divorced from the others. These studies seem to spin in their own orbits, each remaining theoretically or methodologically discrete, profiting little from each other's existence, and failing to feed any useful information to the practitioner conducting daily classroom instruction (Lesser, Fifer & Clark, 1964, p. 148).

Again, Jensen (1969a, p. 117) reflects a similar viewpoint:

> Educational researchers must discover and devise teaching methods that capitalize on existing abilities for the acquisition of those basic skills which students will need in order to get good jobs when they leave school. I believe there will be greater rewards for all concerned if we further explore different types of abilities and modes of learning, and seek to discover how these various abilities can serve the aims of education. This seems more promising than acting as though only one pattern of abilities. . .can succeed educationally, and therefore trying to inculcate this one ability pattern in all children.

Summary of Research Findings

We have seen that the research evidence indicates that members of minority groups may be expected on the average to score less well on most types of objective tests. The cause of the discrepancy, for black examinees at least, does not appear to be the verbal component of the examinations, contrary to popular impression. Further, the validity studies that have been conducted indicate that the usual sort of academic aptitude measure predicts equally well for black and white college undergraduates, but in some instances there has been overprediction of actual grades, or other non-academic types of criteria, instead of the anticipated underprediction.

There is sketchy but provocative evidence to indicate that the atmosphere, both physical and psychological, in which an examination is completed can influence the quality of the performance. The many parameters

of this possibility for group administered measures evidently have not been the subject of research.

Some research is beginning to document the existence of differential patterns of ability in minority groups; this has led to the conviction by some that, instead of searching for minority group members who will fit the traditional patterns of aptitude presently required for the completion of the unit of education, the task should properly be the combined efforts of devising tests that sample a wider range of abilities than are touched on in most academic aptitude tests, and the utilization of these nontraditional aptitudes through an alteration of the techniques of education.

Suggestions for Research

The preceding review of the research literature has made it clear that there is no single study, or series of studies, which will produce the key to the concerns of an admissions testing program. On the other hand, the information already available in the literature can serve a useful function in guiding policy decisions that must be made without the luxury of a time period in which the relevant research can be completed.

An important finding of a general nature, however, has been that the ultimate results of substantive research cannot be anticipated by armchair speculation. In two different problem areas, one concerning the effect of the verbal component in test content, and the other involving the over- or underprediction of criterion performance, the actual results tended to be the reverse of what had been anticipated. Persons in policy-making positions would do well to keep this in mind as the pressures for rapid change increase.

Many of the studies reported here have been conducted on children, and they may therefore be considered of questionable value in this setting; however, the equivalent research on college students does not exist, and evidence perceived to be relevant must be obtained where it can be found. An additional consideration in favor of making use of these studies is that, for the short term at least, many of the problems of educational institutions are similar at all levels. As Hechinger (1970) has said recently, "Educational neglect and inefficiency have made the psychological as well as the academic remedial task the responsibility of higher education." He adds that "the need is to put it back where it belongs"; until this occurs, however, higher education will have more than the usual amounts of common interest with other educational efforts.

Referring again to three possible sources in testing of unfairness to minority group applicants, that is, content, program, and usage, it can be seen that each of the three contain areas in which meaningful research might be pursued.

To consider first test content, the usual aptitude test, if it remains the same, and the curriculum being predicted, if it remains unchanged as well, are likely to continue to yield validities which are roughly the same for both minority and majority group members. However, to the extent that there are changes actually taking place in modern higher education, the past validities are likely to change and alterations in test content will be appropriate. Constant updating of the assurance that the test performance is indicative of scholastic performance is therefore desirable, ideally as a matter of routine. But the nature of the problem is far more complex than can be encompassed by continuing routine validity studies, informative though these might be.

As minority group rights have received more attention, a number of actions have been taken which enormously complicate the once straightforward process of selection. Given this increased attention to minority groups, and our nation's history of race relations, it is improbable that there would be no influence apparent in the treatment and evaluation of these students, and the subsequent alteration of the relationship between current admissions test scores and scholastic performance. There is a subjective element in all academic grading, and racial identity can easily interact with this subjective element in such a way that the validity and apparent usefulness of any given objective measure is altered (Flaugher, Campbell, & Pike, 1969).

An additional complexity concerns the relaxation of the usual admissions standards in order to increase the enrollment of minorities. Given an unchanged curriculum, and no change in the aptitudes demanded by it, there is every reason to expect that these students will be unable to perform well, unless extraordinary efforts are made to motivate students to a degree exceeding that of most of the other students with whom they are competing. On the other hand, if the curriculum demands can be altered to fit the particular abilities existing in these special populations, then the resulting successful performance will once again alter the validity of the selection tests.

The research on these special abilities is just beginning, and it may well be the case that some of the traditional aptitude requirements will remain unaltered in spite of the resourcefulness of the faculty. In the case of those presently measured aptitudes which are determined to be requisite to the completion of the educational unit, the measurement of

these can serve the function of diagnosis, and serve as the focus for subsequent remediation efforts, as Manning (1968) has suggested. At any rate, the role of test content remains crucial in these alterations, and the sort of research that is required must of necessity provide swift feedback and must take cognizance of the changing conditions.

The second possible source of unfairness, designated here as the "program," is defined rather broadly to include the encouragement of minority group members to apply for entrance to the institution and to attempt the examination. This encouragement can be either actual or simply implied by such things as the manner and location in which the test is announced. More directly, it includes the atmosphere in which the test is actually given, and such things as guessing instructions, speededness, attitude of the examiner, and numerous other characteristics of the setting itself. These things are likely to have effects shown by Katz's work to be influential in determining test performance and, as was noted, his studies call attention to the possibility of interactive or cumulative effects of a number of these characteristics simultaneously. Research possibilities are numerous here, seeking to determine which factors beyond the actual content of the test might have intimidating or facilitating influences on minority group applicants.

Finally, in our attention to the aspects of this topic which typically command the attention of research efforts, we must not neglect the third potential source of unfairness, that of the use to which the test scores are put. This problem is not amenable to the usual sorts of research, but the need nevertheless exists for information gathering, perhaps of a field survey sort, to assure that the efforts being made on the other aspects

of the admissions program are not vitiated by such practices as rigid cut-off scores, or requirements of unnecessary and discriminatory levels of particular aptitudes.

References

Anastasi, A. Differential psychology. (3rd ed.) New York: Macmillan, 1959.

Anastasi, A., & Foley, J. P., Jr. Differential psychology. (2nd ed.) New York: Macmillan, 1949.

Blackwood, B. A study of testing in relation to anthropology. Mental Measurements Monograph, 1927, 4, 1-119.

Boney, J. D. Predicting the academic achievement of secondary school Negro students. Personnel and Guidance Journal, 1966, 44, 700-703.

Brown, F. An experimental and critical study of the intelligence of Negro and white kindergarten children. Journal of Genetic Psychology, 1944, 65, 161-175.

Campbell, J. T., Pike, L. W., & Flaugher, R. L. A regression analysis of potential test bias: Predicting job knowledge scores from an aptitude battery. Project Report 69-6. Princeton, N.J.: Educational Testing Service, April 1969.

Cleary, T. A. Test bias: Prediction of grades of Negro and white students in integrated colleges. Journal of Educational Measurement, 1968, 5, 115-124.

Cleary, T. A., & Hilton, T. L. An investigation of item bias. Educational and Psychological Measurement, 1968, 28, 61-75.

Coffman, W. E. Principles of developing tests for the culturally different. In Proceedings of the 1964 Invitational Conference on Testing Problems. Princeton, N.J.: Educational Testing Service, 1965.

Coleman, J., et al. Equality of educational opportunity. U.S. Department of Health, Education, and Welfare, Office of Education. Washington, D.C.: Government Printing Office, 1966. Pp. 242-251.

Cowell, W. R. Special item analysis of the Admission Test for Graduate Study in Business for candidates sponsored by the Consortium for Graduate Study in Business for Negroes. Unpublished manuscript, Educational Testing Service, April 1969.

Davidson, K. S., Gibby, R. G., McNeil, E. B., Segal, S. J., & Silverman, H. A preliminary study of Negro and white differences in Form I of the Wechsler-Bellevue scale. Journal of Consulting Psychology, 1950, 14, 489-492.

De Stephens, W. D. Are criminals morons? Journal of Social Psychology, 1953, 38, 187-199.

Dreger, R. M., & Miller, K. S. Comparative psychological studies of Negroes and whites in the United States. Psychological Bulletin, 1960, 57, 361-402.

Dreger, R. M., & Miller, K. S. Comparative psychological studies of Negroes and whites in the United States: 1959-1965. Psychological Bulletin, 1968 (Monograph Supplement, 70, No. 3, Part 2).

Dubin, J. A., Osburn, H., & Winick, D. M. Speed and practice: Effects on Negro and white test performance. Journal of Applied Psychology, 1969, 53(1), 19-23.

Fifer, G. Social class and cultural group differences in diverse mental abilities. In Proceedings of the 1964 Invitational Conference on Testing Problems. Princeton, N.J.: Educational Testing Service, 1965.

Flaugher, R. L., Campbell, J. T., & Pike, L. W. Ethnic group membership as a moderator of supervisor's ratings. Project Report 69-5. Princeton, N.J.: Educational Testing Service, 1969.

Flaugher, R. L., & Rock, D. A. A multiple moderator approach to the identification of over- and underachievers. Journal of Educational Measurement, 1969, 6(4), 223-228.

Funches, D. L. Correlations between secondary school transcript averages and between ACT scores and grade-point averages of freshmen at Jackson State College. College and University, 1967, 43, 52-54.

Garth, T. R. The problem of racial psychology. Journal of Abnormal and Social Psychology, 1922-23, 17, 215-219.

Grant, D. J., & Bray, D. W. Validation of employment tests for telephone company installation and repair occupations. Journal of Applied Psychology, 1970, 54(1), 7-14.

Hechinger, F. M. The 1970's: Education for what? The New York Times, Monday, January 12, 1970, p. 49.

Higgins, C., & Sivers, C. A comparison of Stanford-Binet and Colored Raven Progressive Matrices IQ's for children with low socioeconomic status. Journal of Consulting Psychology, 1958, 22, 465-468.

Hilgard, E. R. Introduction to psychology. (2nd ed.) New York: Harcourt, Brace, & World, 1957.

Hills, J. R., Klock, J. C., & Lewis, S. Freshman norms for the University System of Georgia, 1960-1962. Atlanta, Georgia: Office of Testing and Guidance, Regents of the University System of Georgia, 1963.

Humphreys, L. G., & Dachler, H. P. Jensen's theory of intelligence. Journal of Educational Psychology, 1969, 60(6), 419-426. (a)

Humphreys, L. G., & Dachler, H. P. Jensen's theory of intelligence: A rebuttal. Journal of Educational Psychology, 1969, 60(6), 432-433. (b)

Iscoe, I., & Pierce-Jones, J. Divergent thinking, age, and intelligence in white and Negro children. Child Development, 1964, 35, 785-797.

Jensen, A. R. How much can we boost IQ and scholastic achievement? Harvard Educational Review, 1969, 39(1), 1-123. (a)

Jensen, A. R. Intelligence, learning ability and socioeconomic status. Journal of Special Education, 1969, 3(1), 23-35. (b)

Jensen, A. R. Jensen's theory of intelligence. Journal of Educational Psychology, 1969, 60(6), 427-431. (c)

Journal of Social Issues. Guidelines for testing minority group children. 1964, 20, 129-145.

Katz, I. Experimental studies of Negro-white relationships. In L. Berkowitz (Ed.), Advances in experimental social psychology, Vol. V. New York: Academic Press, 1970, in press.

Kendrick, S. A. The coming segregation of our selective colleges. College Board Review, Winter 1967-68, 66, 6-13.

Kendrick, S. A., & Thomas, C. L. Transition from school to college. Review of Educational Research, 1970, 40(1), 151-179.

Klein, S. P., Rock, D. A., & Evans, F. R. The use of multiple moderators in academic prediction. Journal of Educational Measurement, 1968, 5, 151-160.

Klineberg, O. Race differences. New York: Harper, 1935.

Klineberg, O. Tests of Negro intelligence. In O. Klineberg (Ed.), Characteristics of the American Negro. New York: Harper, 1944. Pp. 25-96.

Lesser, G., Fifer, G., & Clark, D. Mental abilities of children in different social and cultural groups. Cooperative Research Project, #1635. Washington, D.C.: Office of Education, U.S. Department of Health, Education and Welfare, 1964.

Linn, R. L., & Werts, C. E. Considerations for studies of test bias. Growth Study Paper Number 53. Unpublished manuscript, Educational Testing Service, 1970.

Manning, W. H. The measurement of intellectual capacity and performance. Journal of Negro Education, 1968, 37(3), 258-267.

McKelpin, J. P. Some implications of the intellectual characteristics of freshmen entering a liberal arts college. Journal of Educational Measurement, 1965, 2, 161-166.

Moore, C. L., Jr., & MacNaughton, J. F. An exploratory investigation of ethnic differences within an industrial selection battery. Paper presented at the meeting of the American Psychological Association, New York, 1966.

Morgan, L. B. The calculated risks--a study of success. College and University, 1968, 43, 203-206.

Munday, L. A. Predicting college grades in predominantly Negro colleges. Journal of Educational Measurement, 1965, 2, 157-160.

Olsen, M. Summary of main findings on the validity of the CEEB tests of developed ability as predictors of college grades. Statistical Report 57-14. Princeton, N.J.: Educational Testing Service, 1957.

Perlberg, A. Predicting academic achievements of engineering and science college students. Journal of Educational Measurement, 1967, 4, 241-246.

Peterson, R. E. Predictive validity of a brief test of academic aptitude. Educational and Psychological Measurement, 1968, 28, 441-444.

Pettigrew, T. F. A profile of the Negro American. Princeton, N.J.: Van Nostrand, 1964.

Pressey, S. L., & Teter, G. F. A comparison of colored and white children by means of a group scale of intelligence. Journal of Applied Psychology, 1919, 3, 277-282.

Roberts, S. O. Studies in identification of college potential. Nashville, Tenn.: Fisk University, Department of Psychology, 1962. (Mimeo.)

Roberts, S. O. Comparative validity study of CEEB and CIEP test programs. Nashville, Tenn.: Fisk University, Department of Psychology, 1964. (Mimeo.)

Rock, D. A. Student characteristics as moderators within curriculum. Paper presented at the meeting of the American Psychological Association, San Francisco, 1968.

Rock, D. A. Motivation, moderators, and test bias. Toledo Law Review, 1970, in press.

Sattler, J. M. Racial "experimenter effects" in experimentation, testing, interviewing, and psychotherapy. Psychological Bulletin, 1970, 73(2), 137-160.

Semler, I. J., & Iscoe, I. Structure of intelligence in Negro and white children. Journal of Educational Psychology, 1966, 57, 326-336.

Shuey, A. The testing of Negro intelligence. (2nd ed.) New York: Social Science Press, 1966.

Stanley, J. C., & Porter, A. C. Correlation of scholastic aptitude test scores with college grades for Negroes versus whites. Journal of Educational Measurement, 1967, 4, 199-218.

Stodolsky, S. S., & Lesser, G. Learning patterns in the disadvantaged. Harvard Educational Review, 1967, 37(4), 546-593.

Strong, A. C. Three hundred fifty white and colored children measured by the Binet-Simon measuring scale of intelligence: A comparative study. Pedagogical Seminary, 1913, 20, 485-515.

Tenopyr, M. L. Race and socioeconomic status as moderators in predicting machine-shop training success. A paper presented in a symposium on "Selection of Minority and Disadvantaged Personnel" at the meeting of the American Psychological Association, Washington, D.C., 1967.

Thomas, C. L., & Stanley, J. C. The effectiveness of high school grades for predicting college grades of Negro students: An exploratory study. New York: Teachers College, Columbia University, 1969. (Mimeo.)

Time, April 6, 1970, 95(14). Getting it together: The young blacks. (Situation Report, p. 46.)

Vernon, P. E. Ability factors and environmental influences. American Psychologist, 1965, 20, 723-733.

Footnotes

[1]An earlier version of this paper was prepared for the Research and Development Committee, Admission Test for Graduate Study in Business.

[2]The author is indebted to Joel T. Campbell, Irwin Katz, Winton H. Manning and John A. Winterbottom for reviewing earlier versions of this paper.